

TInCAP – ein interdisziplinäres Korpus zu Ambiguitätsphänomenen

Hartmann, Jutta

jutta.hartmann@uni-tuebingen.de
Universität Tübingen, Deutschland/ Universität Stuttgart,
Deutschland

Sauter, Corinna

corinna.sauter@uni-tuebingen.de
Universität Tübingen, Deutschland

Schole, Gesa

gesa.schole@uni-tuebingen.de
Universität Tübingen, Deutschland

Wagner, Wiltrud

wiltrud.wagner@uni-tuebingen.de
Universität Tübingen, Deutschland

Gietz, Peter

peter.gietz@daasi.de
DAASI International

Winkler, Susanne

susanne.winkler@uni-tuebingen.de
Universität Tübingen, Deutschland

Ambiguitäten allerorten: Ambiguität ist ein integraler Bestandteil menschlicher Kommunikation. Sie kann unbeabsichtigt produziert werden, wie in (1a), oder zu strategischen Zwecken eingesetzt werden, z. B. für komische Effekte wie in (1b).

1a. William isn't drinking because he's unhappy (vgl. Hirschberger / Avesani 1997).

1b. ...men who can shear sheep and women with long hair... (vgl. Cutting from BBC News Website).

Ambiguität findet sich vornehmlich in sprachlichen Ausdrücken, kann jedoch auch in der Interaktion mit Bildern oder in Bildern selbst zu finden sein, sowie auf nicht-sprachliche Kommunikation übertragen werden. Daher eröffnet das Thema Ambiguität ein interdisziplinäres Forschungsfeld, an dem neben Sprach- und Literaturwissenschaft auch Rhetorik, Psychologie, Theologie, Rechtswissenschaft und Medienwissenschaften größtes Interesse bekunden (siehe beispielsweise Klein / Winkler 2010; Winkler 2015).

Das Datenbankprojekt TInCAP (Tübingen Interdisciplinary Corpus of Ambiguity Phenomena), das im Rahmen des interdisziplinären Graduiertenkollegs

1808 „Ambiguität: Produktion und Rezeption“ entsteht, zielt darauf ab, Belege von Ambiguität verschiedener Provenienz zu sammeln und diese interdisziplinär zu annotieren und nachhaltig zu speichern. Dabei stehen drei Ziele im Vordergrund: (a) die interdisziplinäre Auseinandersetzung mit dem Thema Ambiguität durch die Erstellung eines gemeinsamen Annotationsschemas, (b) die Nachhaltigkeit der gesammelten Daten und (c) die Zugänglichkeit der Datensammlung für die nationale und internationale Forschungsgemeinschaft.

ANNOTATION. Jeder Beleg zur Ambiguität kann interdisziplinär im Hinblick auf fünf verschiedene Aspekte annotiert werden. Die (i) Kommunikationsebene legt fest auf welcher Ebene die Ambiguität in der Kommunikation eine Rolle spielt, beispielsweise auf der Ebene der fiktiven Charaktere vs. Erzähler-Leser vs. konkrete Kommunikation. Wir unterscheiden außerdem (ii) zwischen strategischem vs. nicht-strategischem Einsatz der Ambiguität in Produktion und Rezeption. Darüber hinaus wird (iii) sowohl die Ebene des Auslösers der Ambiguität (z. B. auf Wortebene, Phrasenebene etc.) annotiert als auch ihre Reichweite, d. h. bis zu welcher Ebene sie für die Interpretation relevant ist. Als weiteren Punkt kennzeichnet (iv) eine qualitative Annotation wie sich die unterschiedlichen Lesarten zueinander verhalten: Sind sie voneinander abgeleitet oder komplett unabhängig? Spielt Vagheit eine Rolle? Nicht zuletzt sieht die Datenbank die Möglichkeit vor, (v) disziplininterne Begrifflichkeiten zur Beschreibung des behandelten Phänomens zu verwenden. Dieses interdisziplinär erarbeitete und anwendbare Schema erlaubt es uns langfristig Korrelationen in den Daten zu finden, die über die unterschiedlichen Modi (unterschiedliche Typen von Ambiguität in unterschiedlichen Texttypen / Bildern) hinweg gelten, sodass wir damit ein genuin interdisziplinäres Forschungsergebnis erreichen können.

NACHHALTIGKEIT. Ein wichtiges Ziel der Datenbank ist es, die gesammelten Daten langfristig und nachhaltig zu speichern. Dazu haben wir ein XML-Schema entwickelt, das weitestgehend TEI-konform [5] ist. Diese XML-Dateien können im Rahmen der universitären Infrastruktur langfristig gespeichert, katalogisiert und zugänglich gemacht werden. Bei Video-, Audio- und Bilddateien halten wir uns an die üblichen Standards für nachhaltige Datenformate.

INTERFACE. Für die aktive Arbeitsphase mit der Datenbank im Rahmen des GRK 1808 und für die Zugänglichkeit für die (inter)nationale Forschungsgemeinschaft haben wir eine Datenbankanwendung spezifiziert, die von einem externen Dienstleister implementiert wird. Dabei setzen wir auf die objektorientierte hierarchische Datenbanktechnologie LDAP (vgl. Zeilenga 2006), die bereits im BMBF-Projekt RiR eingesetzt wurde. So lässt sich nicht nur die XML-Hierarchie bestens abbilden, sondern es wird auch eine sichere und feingranulare Zugriffskontrolle ermöglicht. Mittels einer entsprechend angepassten Synchronisierungssoftware konnte die

Datenbank während der XML-Erfassungsphase ständig aktualisiert werden. Eine webbasierte Benutzeroberfläche ermöglicht u.a. komplexe Suchen in den verschiedenen Hierarchieebenen, wobei mehrere Einträge in der Hierarchie (also z. B. ein Haupteintrag sowie mehrere Annotationseinträge und bibliographische Einträge) als ein virtueller Eintrag zusammengezogen werden. Über die Benutzeroberfläche können neue Einträge erstellt und vorhandene Einträge modifiziert werden, wobei die Zugriffskontrolle erlaubt, auch nur Teile eines solchen virtuellen Eintrags sichtbar/bearbeitbar zu machen. Gleichzeitig erlaubt der Export einzelner bzw. aller Datensätze, im XML-Format die Nachhaltigkeit der eingegebenen Daten auch über einen längeren Zeitraum hinweg sicherzustellen.

Damit zeigt das Datenbankprojekt, wie sich interdisziplinäre inhaltliche Arbeit innovativ mit den Zielen der Nachhaltigkeit verknüpfen lässt, ohne die Benutzerfreundlichkeit in der aktiven Arbeitsphase zu vernachlässigen.

Bibliographie

Cutting from BBC News Website, quoted BBC 4, Friday Night Comedy, the News Quiz, Series 82, Episode 2; Broadcasted: 15.Nov 2013.

Hirschberg, Julia / Avesani, Cinzia (1997): „The role of prosody in disambiguating potentially ambiguous utterances in English and Italian“, in: Botinis, Antonis / Kouroupetroglou, Georgios / Carayannis, George (eds.): *Intonation. Theory, Models and Applications* 189–192.

Klein, Wolfgang / Winkler, Susanne (eds.) (2010): *Ambiguität. Zeitschrift für Literaturwissenschaft und Linguistik* 40, 158. Stuttgart: Metzler.

RiR (2012-2015): *Relationen im Raum* <http://www.steinheim-institut.de/wiki/index.php/RiR> [letzter Zugriff 15. Februar 2016].

TEI Consortium (eds.): *Guidelines for Electronic Text Encoding and Interchange* <http://www.tei-c.org/P5/> [letzter Zugriff 15. Februar 2016].

Winkler, Susanne (ed.) (2015): *Ambiguity. Language and Communication*. Berlin / New York: de Gruyter.

Zeilenga, Kurt (2006): *Lightweight Directory Access Protocol (LDAP)*. Directory Information Models, IETF RFC 4512, June 2006.